An Algorithm Based on Singular Spectrum Analysis for Change-Point Detection

Valentina Moskvina and Anatoly Zhigljavsky*

School of Mathematics, Cardiff University,
Cardiff, UK

ABSTRACT

This paper is devoted to application of the singular-spectrum analysis to sequential detection of changes in time series. An algorithm of change-point detection in time series, based on sequential application of the singular-spectrum analysis is developed and studied. The algorithm is applied to different data sets and extensively studied numerically. For specific models, several numerical approximations to the error probabilities and the power function of the algorithm are obtained. Numerical comparisons with other methods are given.

Key Words: Change-point detection; Singular-spectrum analysis; Sequential algorithm; Singular value decomposition.

*Correspondence: Anatoly Zhigljavsky, School of Mathematics, Cardiff University, Senghennydd Rd., Cardiff, UK CF24 4YH; Fax: +44(0) 29 20 8741 99; E-mail: zhigljavskyaa@cardiff.ac.uk.

319

1. INTRODUCTION

Detection of changes in time series constitutes a very important branch of applied statistics. There are many methods available (an excellent introduction to the subject is the monograph Basseville and Nikiforov (1993)), but the problem is by no means solved. For example, the majority of the change-point detection techniques developed in the past is parametric and based on simple models. The method we develop and study is non-parametric and can be applied to analyze time series of complex structure.

The method is based on sequential application of the so-called Singular-Spectrum Analysis (SSA), which is a very powerful but relatively unknown technique of time series analysis (see Sec. 2). SSA incorporates the elements of classical time series analysis, multivariate statistics, multivariate geometry, dynamical systems and signal processing, see Broomhead and King (1986), Danilov and Zhigljavsky (1997), Elsner and Tsonis (1996), Goljandina et al. (2001), Vautard et al. (1992), and references therein. The main idea of SSA is performing a singular value decomposition (SVD) of the trajectory matrix obtained from the original time series with a subsequent reconstruction of the series. A version of SSA sufficient for our aims is described in Sec. 2.

In Sec. 3 the change-point detection algorithm is presented. This algorithm is based on the idea that if at a certain time moment τ the mechanism generating the time series x_t has changed, then an increase in the distance between a subspace of $\mathbb{R}^M$ spanned by certain eigenvectors of the so–called lag–covariance matrix, and M-lagged vectors $(x_{j+1}, \ldots, x_{j+M})$ is to be expected for $j \geq \tau$.

Section 4 briefly describes the ways of choosing the parameters of the proposed algorithm. There are four parameters in the algorithm in addition to the two SSA parameters. One of them is the window length N, the parameter defining the moving window where the SSA decomposition is performed. Two parameters define the test sample which closeness to the subspace is tested at each iteration of the algorithm (there are relatively clear rules for choosing these parameters). The fourth parameter is the threshold h which is usually chosen to fix the significance level (alternatively, the average run length) at some level.

In Sec. 5 three examples are considered to illustrate the basic features of the proposed scheme. No comparison with other change-point algorithms is made in this section; this comparison is delayed until Sec. 7.

In Secs. 6 and 7 results of simulations are reported that are used to assess the quality of different approximations and compare the proposed algorithm against some classical change-point detection algorithms.

The main detection statistic in the proposed algorithm can be represented as a moving sum of the squared distances to the l-dimensional subspace that describes the main structure of the series; it can often be associated with the signal. Under the assumption that the signal is well described by this subspace and there is no change in the structure of the series, the residuals are considered as i.i.d.r.v.

In Sec. 6 the behaviour of the error probability of type I is numerically studied and compared against the test when the detection statistic is the moving weighted sum of normal random variables. The results show that for suitably large values of M and T the error probabilities are reasonably close.

In Sec. 7.1 the behaviour of the error probability of type II is numerically studied for the hypothesis of a change in mean of a normal distribution and the main test, when the detection statistic is the moving weighted sum of squares. It is demonstrated that the algorithm shows the best performance when $Q = 1$ and M is slightly smaller than T.

In Sec. 7.2, for the same problem, the error probabilities of type I and II are displayed for a version of the CUSUM algorithm (which is optimum for this problem in some other sense); the power of the proposed algorithm is compared with the power of CUSUM. It is found that the proposed algorithm is approximately three times worse than CUSUM: it needs three times less observations for CUSUM to detect the change after it happened than for the proposed algorithm.

The situation changes dramatically when the change is in variance of the normally distributed variables. In this case the proposed algorithm becomes six or more times better than the version of CUSUM we consider. This is discussed in Sec. 7.3. In many practical problems the changes in time series reveal in both mean and variance of the series, so a reasonable change-point detection algorithm should perform well in both situations.

In Sec. 8 we study the error probabilities of the proposed algorithm for the problem of detecting a change in mean and separately, variance of normal r.v. by fixing n (instead of considering a moving sum we thus consider only one time moment). Of course, this assumption significantly simplifies the problem, and we are able to obtain exact formulae for the error probabilities. These probabilities are compared against the error probabilities for the tests based on using the sum of r.v. and absolute value of sum of r.v. The conclusion about the relative performance of the tests is very similar to that derived in Sec. 7.

A C++ program for change-point detection with graphical representation of results is written. The program can be downloaded from the web-site

http://www.cf.ac.uk/maths/stats/changepoint/

For a comprehensive description of commands, see Help in this program.

The algorithm described in the present article is a further elaboration of the SSA-based change-point detection methodology developed in Goljandina et al. (2001) and, especially, in Moskvina (2001), and Moskvina and Zhigljavsky (2000). We also refer to these references for a further discussion of the methodological and statistical issues and to a discussion on the comparison of the present methodology with the variety of existing change-point detection algorithms and programs.

2. SINGULAR-SPECTRUM ANALYSIS

2.1. Description of the Algorithm

The basic version of SSA consists of four steps, which are performed as follows. (See Goljandina et al. (2001) for a more general version of SSA and various modifications of the basic scheme.)

Step 1. Embedding

Let $x_1, x_2, \ldots, x_N$ be a time series, M ($M \leq N/2$) be some integer called 'lag' and let $K = N - M + 1$. Define the matrix

$$\mathbf{X} = (x_{ij})_{ij=1}^{M,K} = \begin{pmatrix} x_1 & x_2 & x_3 & \ldots & x_K \\ x_2 & x_3 & x_4 & \ldots & x_{K+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_M & x_{M+1} & x_{M+2} & \ldots & x_N \end{pmatrix} \tag{1}$$

and call it the trajectory matrix. Obviously $x_{ij} = x_{i+j-1}$ so that the matrix $\mathbf{X}$ has equal elements on the diagonals $i + j = \text{const}$. In terms of the matrix theory, $\mathbf{X}$ is a Hankel matrix.

Consider $\mathbf{X}$ as a multivariate data with M characteristics and $K = N - M + 1$ observations and let $\mathbf{R} = \mathbf{X}\mathbf{X}^T$. (We shall call $\mathbf{R}$ the lag–covariance matrix.) Denote the columns of $\mathbf{X}$ by $X_1, X_2, \ldots, X_K$. These vectors lie in an M-dimensional space $\mathbb{R}^M$.

Step 2. Singular Value Decomposition

Singular value decomposition (SVD) of $\mathbf{R} = \mathbf{X}\mathbf{X}^T$ provides us with the collections of M eigen–values, eigen–vectors and principal components.

Denote by $\lambda_1, \ldots, \lambda_M$ the eigenvalues of $\mathbf{R}$ and assume that they are arranged in the decreasing order, so that $\lambda_1 \geq \cdots \geq \lambda_M \geq 0$. Let $U_1, \ldots, U_M$ be the corresponding orthonormal eigenvectors of $\mathbf{R}$ (in SSA literature they are often called "empirical orthogonal functions" or simply EOFs). Let also d be the number of nonzero eigenvalues λ_i and V_i be the eigenvectors of the matrix $\mathbf{X}^T\mathbf{X}$ (these vectors are sometimes called principal components). We have $V_i = \mathbf{X}^T U_i$ for $i = 1, \ldots, d$. As a result of the SVD we obtain a representation $\mathbf{X} = \mathbf{X}_1 + \cdots + \mathbf{X}_d$, where $\mathbf{X}_i = \sqrt{\lambda_i} U_i V_i^T$ are rank-one biorthogonal matrices.

Step 3. Grouping

Split the set of indices $\{1, 2, \ldots, d\}$ into two groups, namely

$$I = \{i_1, \ldots, i_l\} \quad \text{and} \quad \bar{I} = \{1, \ldots, d\} \setminus I \tag{2}$$

and sum the matrices $\mathbf{X}_i$ within each group. The result of the step is the representation

$$\mathbf{X} = \mathbf{X}_I + \mathbf{X}_{\bar{I}}, \quad \text{where} \quad \mathbf{X}_I = \sum_{i \in I} \mathbf{X}_i \quad \text{and} \quad \mathbf{X}_{\bar{I}} = \sum_{i \notin I} \mathbf{X}_i. \tag{3}$$

Step 4. Reconstruction

Perform the averaging over the diagonals $i + j = \text{const}$ of the matrices $\mathbf{X}_I$ and $\mathbf{X}_{\bar{I}}$. Applying then twice the one-to-one correspondence between the series of length N and the Hankel matrices of size $M \times K$, we obtain two series (denote them by $\{z_t\}$ and $\{\varepsilon_t\}$) and the SSA decomposition of the original series; that is, a decomposition of $\{x_t\}$ into a sum of two series

$$x_t = z_t + \varepsilon_t. \tag{4}$$

Here the series $\{z_t\}$ (obtained from the diagonal averaging of $\mathbf{X}_I$) can often be associated with signal and the residual series $\{\varepsilon_t\}$ with noise.

3. CHANGE-POINT DETECTION

3.1. The Main Idea

The selection of the group I of $l \leq \text{rank}\mathbf{X}$ indices on the third step of the basic SSA algorithm implies the selection of l eigen-vectors, which

determine an l-dimensional subspace in the M-dimensional space $\mathbb{R}^M$ of vectors X_j.

The distance between the vectors X_j $(j = 1, \ldots, K)$ and this l-dimensional subspace can be reduced to rather small values (it is controlled by the choice of I). It should stay reasonably small for $X_j, j > K$, if the time series $\{x_t\}_{t=1}^N$ is continued for $t > N$ and there is no change in the mechanism generating x_t.

The proposed change-point detection algorithm is based on the observation that if at a certain time moment $N + \tau$ the mechanism generating x_t $(t \geq N + \tau)$ has changed, then an increase in the distance between the l-dimensional subspace and vectors X_j for $j \geq K + \tau$ is to be expected.

To make the change-point detection algorithm sequential and to accommodate it to a slow change in the time series structure, to outliers and to the case of multiply changes, for each n we apply SVD to the trajectory matrix computed in a time interval $[n + 1, n + N]$ of length N rather than to the trajectory matrix (1). Here n is the iteration number and N is the length of the time interval where the trajectory matrix is computed.

3.2. Description of the Algorithm

Let $x_1, x_2, \ldots$ be a time series and N, M, l, p and q be fixed integers so that $l < M \leq N/2$ and $0 \leq p < q$. For each $n = 0, 1, \ldots$ we execute the following operations.

Stage 1. Construction of an l-Dimensional Space

At this stage we perform the first three steps of the SSA algorithm in the time interval $[n + 1, n + N]$.

1. Construct the trajectory matrix

$$\mathbf{X}_B^{(n)} = \begin{pmatrix} x_{n+1} & x_{n+2} & x_{n+3} & \cdots & x_{n+K} \\ x_{n+2} & x_{n+3} & x_{n+4} & \cdots & x_{n+K+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{n+M} & x_{n+M+1} & x_{n+M+2} & \cdots & x_{n+N} \end{pmatrix} \tag{5}$$

where $K = N - M + 1$ (in the change–point detection algorithm these matrices are called *base matrices*). The columns of the matrix $\mathbf{X}_B^{(n)}$ are the vectors

$$X_j^{(n)} = (x_{n+j}, \ldots, x_{n+j+M-1})^T \quad \text{with} \quad j = 1, \ldots, K.$$

2. Perform the SVD of the lag-covariance matrix $\mathbf{R}_n = \mathbf{X}^{(n)}(\mathbf{X}^{(n)})^T$. This gives us a collection of M eigenvectors.
3. Select a particular group I of $l < M$ of these eigenvectors; this determines an l-dimensional subspace $\mathcal{L}_{n,I}$ of the M-dimensional space $\mathbb{R}^M$.

Stage 2. Construction of the Test Matrix

Construct the matrix $\mathbf{X}_T^{(n)}$ of size $M \times Q$, whose columns are the vectors $X_j^{(n)}(j = p+1, \ldots, p+Q)$; that is,

$$\mathbf{X}_T^{(n)} = \begin{pmatrix} x_{n+p+1} & x_{n+p+2} & x_{n+p+3} & \cdots & x_{n+q} \\ x_{n+p+2} & x_{n+p+3} & x_{n+p+4} & \cdots & x_{n+q+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{n+p+M} & x_{n+p+M+1} & x_{n+p+M+2} & \cdots & x_{n+q+M-1}, \end{pmatrix}$$

where $q = p + Q$. This matrix is called *test matrix*.

Stage 3. Computation of the Detection Statistics

The detection statistics are:

- $\mathcal{D}_{n,I,p,q}$, the sum of squared Euclidean distances between the vectors $X_j^{(n)}(j = p+1, \ldots, q)$ and the $l-$ dimensional subspace $\mathcal{L}_{n,I}$ of $\mathbb{R}^M$ (see Eq. (8) for the explicit formula of $\mathcal{D}_{n,I,p,q}$).
- $S_n = \tilde{\mathcal{D}}_{n,I,p,q}/\mu_{n,I}$, the normalized sum of squares of distances. Here

$$\tilde{\mathcal{D}}_{n,I,p,q} = \frac{1}{MQ} \mathcal{D}_{n,I,p,q} \tag{6}$$

(the sum of squared distances is normalized to the number of elements in the test matrix) and $\mu_{n,I}$ is an estimator of the normalized sum of squared distances $\tilde{\mathcal{D}}_{j,l,p,q}$ at the time intervals $[j+1, j+m]$ where the hypothesis of no change can be accepted. We use $\mu_{n,I} = \tilde{\mathcal{D}}_{m,I,0,K}$ where m is the largest value of $m \le n$ so that the hypothesis of no change is accepted.

- The CUSUM-type statistic

$$W_1 = S_1, \qquad W_{n+1} = (W_n + S_{n+1} - S_n - \kappa/\sqrt{MQ})^+, \quad n \geq 1,$$

where $(a)^+ = \max\{0, a\}$ for any $a \in \mathbb{R}$ and κ is a small nonnegative constant (shift down in case of the null hypothesis $\mathbf{H_0}$), a reasonable value of κ is $\kappa = 1/(3\sqrt{MQ})$, see Moskvina (2001).

3.3. The Decision Rule

Large values of $\mathcal{D}_{n,I,p,q}$, S_n and W_n indicate on a change in the structure of the time series. The first point with non-zero value of W_n before this statistic has reached a high value should be considered as an estimator of the change-point. The algorithm announces the structural change if for some n we observe $W_n > h$ with the threshold

$$h = \frac{2t_\alpha}{MQ}\sqrt{\frac{1}{3}Q(3MQ - Q^2 + 1)}, \tag{7}$$

where t_α is the $(1-\alpha)$-quantile of the standard normal distribution (see Sec. 8.1 and formulae (23)–(24)). Under the assumptions of the asymptotic normality of $\mathcal{D}_{n,I,p,q}$ (this holds when the window width N and the lag M are large enough) and that the time series is a sum of a signal (which is recovered by the SSA decomposition) and white noise, for this decision rule the probability of a false alarm for a fixed n is approximately 0.001 (see Moskvina (2001) and Moskvina and Zhigljavsky (2000)) for the statistical aspects of the procedures of this kind).

3.4. Computational Details

The way of constructing the base and test matrices is illustrated in Fig. 1.

Consider computation of the distances between the vectors $X_j^{(n)}$ and the space $\mathcal{L}_{n,I}$. Let the group I be $I = \{i_1, \ldots, i_l\}$. Then the space $\mathcal{L}_{n,I}$ is spanned by the eigenvectors $U_{i_1}, \ldots, U_{i_l}$. Since the eigenvectors are orthonormal, the square of the Euclidean distance between an M-vector Z and the subspace $\mathcal{L}_{n,I}$ spanned by the l eigenvectors $U_{i_1}, \ldots, U_{i_l}$, is just

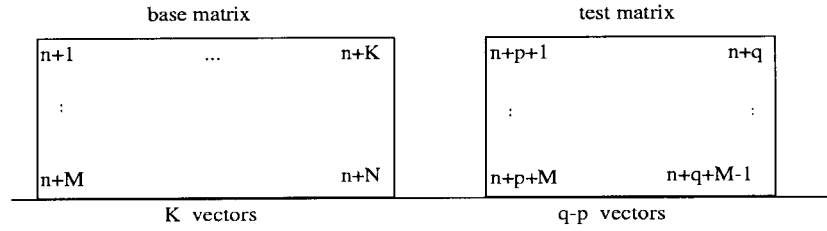$$\|Z\|^2 - \|U^T Z\|^2 = Z^T Z - Z^T U U^T Z$$

Figure 1. Construction of the base and test matrices.

where $\| \cdot \|$ is the usual Euclidean norm and U is the $M \times l$-matrix with the columns $U_{i_1}, \ldots, U_{i_l}$. Therefore

$$\mathcal{D}_{n, I, p, q} = \sum_{j=p+1}^{q} ((X_j^{(n)})^T X_j^{(n)} - (X_j^{(n)})^T U U^T X_j^{(n)}). \tag{8}$$

4. CHOICE OF PARAMETERS

Significant changes in time series structure will be detected for any reasonable choice of parameters. To detect small changes in noisy series a tuning of parameters may be required. Below you can find some recommendations concerning this tuning.

4.1. Parameters of the SSA Algorithm: Lag M and Group I

The parameters of the SSA algorithm are the lag M and the group I. To choose them, we have to follow standard SSA recommendations. (For an extensive discussion of this problem we refer to Goljandina et al. (2001).) The separability characteristics (including the **w**-correlation) play a very important role here. One of the separability characteristics is the so-called **w**-correlation which is defined as

$$\rho = \frac{\sum_{t=1}^{N} w_t z_t \varepsilon_t}{\sqrt{\sum_{t=1}^{N} w_t z_t^2 \sum_{t=1}^{N} w_t \varepsilon_t^2}}, \tag{9}$$

where w_t are the weights such that $w_t = w_{M, 0, K}(t)$ with the function $w_{M, 0, K}(t)$ defined in Eq. (22).

In an ideal situation, the components in Eq. (4) must be "independent." Achieving "independence" (or "separability") of the components z_t and ε_t in the SSA decomposition Eq. (4) is of prime importance in SSA.

If N is not very large, which should be regarded as the most interesting case in practice, the main recommendation is to choose $\lfloor M = N/2 \rfloor$ and $I = \{1, \ldots, l\}$, where l is such that the first l components provide a good description of the signal and the lower $M - l$ components correspond to noise.

To choose l, it is advised to make a visual inspection of the SSA decomposition of the whole series and some large parts of the series before applying the change-point detection algorithm. Alternatively, if the problem is really sequential and a preliminary study of the time series is not possible, then the recommendation is to use all the visual SSA tools in the first part of the series to choose l.

If l is too small (underfitting), then we miss a part of the signal and therefore we can miss a change (which may occur in the underestimated components). Alternatively, if l is too large (overfitting), then we approximate a part of noise together with the signal; therefore, finding a change in signal becomes more difficult.

4.2. Length and Location of the Test Sample: p, q

A general recommendation is to choose $p \geq K$. In this case the columns of the base and test matrices are generally different. In this case the algorithm is more sensitive to changes than its more economical version (in the sense of the number of x_t involved at each iteration) when $p < K$ and thus some of the columns of the base and test matrices coincide.

To get a smooth behaviour of the test statistics $\mathcal{D}_{n,I,p,q}$ we need to select q slightly larger than p. If the difference $q - p$ is too large, then the behaviour of $\mathcal{D}_{n,I,p,q}$ becomes too smooth; this happens, for example, when $p = 0$ and $q = K$ (that is, the base and test matrices coincide).

4.3. Window Width N

The choice of N depends on what kind of structural changes we are looking for. If we allow small gradual changes in the time series then we could not take N very large. Also, if N is too large, then we could either miss or smooth out all the changes in our time series. Alternatively, if we take N small, then an outlier could be recognized as a structural change

(concerning detecting the outliers, see Example 3 of Sec. 5). A general rule is to choose N reasonably large.

4.4. Centering

If the trend of original series $\{x_t\}$ contains either a constant or a linear component which influence we would prefer to neglect, then it can be worthwhile to make centering of the base and test matrices (that is, to subtract row and perhaps column averages from the elements of the matrices). The effect of this operations on the behaviour of the quadratic form $\mathcal{D}_{n,I,p,q}$ is studied in Moskvina (2000).

5. NUMERICAL EXAMPLES

In this section we present numerical examples illustrating the behavior of the procedure (many more examples can be found in Moskvina (2001)). Note, that in the graphs we plot the normalized detection statistics (6) shifted on $q + \tilde{M} - 1$ points to the right to align the point of the first increase of these statistics and the change-point. Thus, we plot the values of

$$\tilde{\mathcal{D}}_{n',p,q} = \frac{1}{MQ} \mathcal{D}_{n+M+q-1,\,p,\,q}.$$

Example 1. (*Airlines data*, Fig. 2.)

This celebrated data, see for example, Box and Jenkins (1970), give logarithms of monthly totals (in thousands) of international airline passengers for January 1949–December 1960. There are only 144 data points so we have selected a rather small N, namely $N = 36$. (For the sake of precision of SSA decompositions N has to be proportional to the main period which is 12.) We have also taken $M = N/2 = 18$ and $p = 18$, according to the recommendations of Sec. 4, and $q = 30$ (value $q = 36$ would be a little too large: there is not enough data). To choose l, we have made the standard SSA decomposition of the whole series with $M = 18$. The results of the decomposition are displayed in Fig. 2 (a): the main trend is described by the 1st and 6th eigenvectors (principal components), the main period (12-months) is described by the 2nd and 3rd components and the second main period (6-months) is represented by the 4th and 5th principal components. (The series reconstructed from these two components is
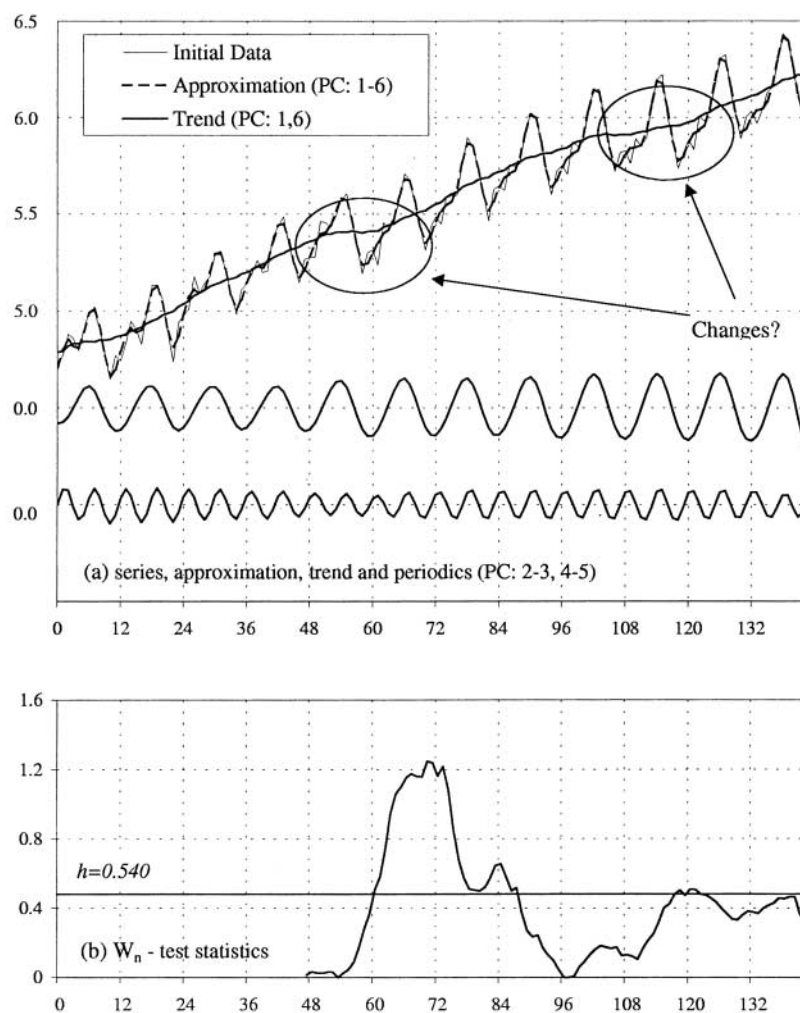
Figure 2. Logarithm of airline passenger numbers. Parameters: $N = 36$, $M = 18$, $Q = 12$, PC: 1–6.

shifted down, which results in the second zero in the plot in Fig. 2(a).)
Figure 2(b) shows that the test statistic clearly indicates on two time intervals where changes in trend have possibly occurred (alternatively, this may indicate an additional five-year cycle in the data). The threshold h is computed according to Eq. (7) with $t_\alpha = 1.2815$ (this corresponds to 0.9-quantile of the standard normal distribution), its value is $h = 0.540$.

Example 2. (*A change in a correlation structure*, Fig. 3.)

In this example we have two-dimensional series $\{e_t^{(1)}, e_t^{(2)}\}$ with independent vectors

$$\begin{pmatrix} e_t^{(1)} \\ e_t^{(2)} \end{pmatrix} \sim \begin{cases} N(0, \sigma^2 I_2) & \text{for } t = 1, \ldots, 200 \\ N(0, \sigma^2 \Sigma) & \text{for } t = 201, \ldots, 400, \end{cases} \tag{10}$$

where $\sigma = 1$,

$$I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}.$$

The individual series $e_t^{(j)}$ ($j = 1, 2$) do not have changes, see Fig. 3(a,b); the change occurs in the correlation structure of the series. To detect this change, we consider the sum $e_t' = e_t^{(1)} + e_t^{(2)}$ (see Fig. 3(c)) and test a change in it.

The modified CUSUM test

$$g_k' = \sum_{i=1}^{k} (e_i^{(1)} + e_i^{(2)})$$

and normalized moving sum test with $m = 100$

$$\tilde{g}_k' = \frac{1}{m} \sum_{i=k+1}^{k+m} \left(e_i^{(1)} + e_i^{(2)} \right)$$

do not reflect the change, see Fig. 3(d,e). However, the normalized moving sum of squares

$$g_k = \frac{1}{2m\sigma^2} \sum_{i=k+1}^{k+m} \left(e_i^{(1)} + e_i^{(2)} \right)^2$$

($m = 100$) does reflect the change, see Fig. 3(f). Note that this algorithm corresponds to the version $p = N$ and $q = N + 1$ with $M = 100$, of our algorithm.

The result of this example can be explained as follows. The change-point model Eq. (10) is reduced to a change in variance for the series $e_i^{(1)} + e_i^{(2)}$ ($i = 1, \ldots, 400$) which is a sequence of independent normal r.v. with zero mean and variances $2\sigma^2$ for $i \leq 200$ and $3\sigma^2$ for $i > 200$. It is well-known, however (see e.g., Basseville and Nikiforov (1993)), that the likelihood ratio statistics for this problem is the sum of squares of $(e_i^{(1)} + e_i^{(2)})^2$.

Example 3. (*Presence of an outlier*, Figs. 4–5.)

Assume that we have an outlier at a point τ and the base and test matrices in the change-point detection algorithm do not intersect
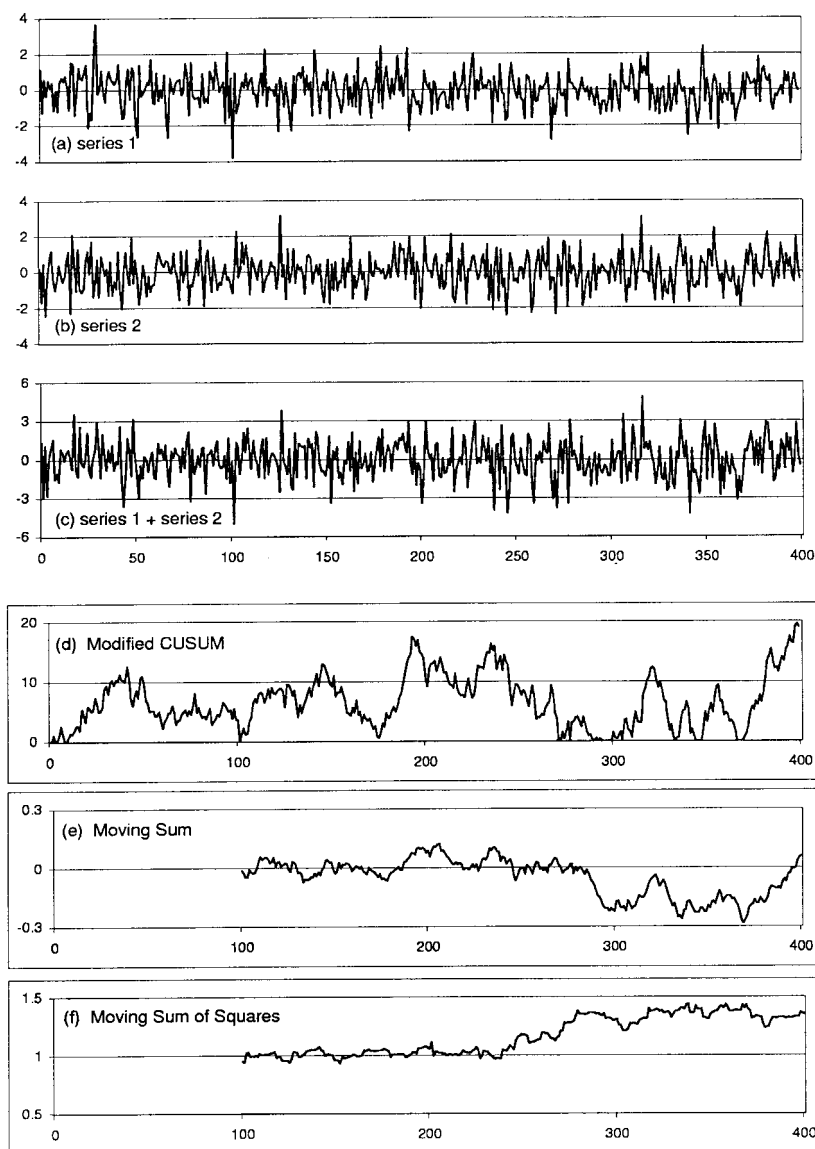
Figure 3. Model of Example 2.

MARCEL DEKKER, INC. • 270 MADISON AVENUE • NEW YORK, NY 10016

(that is, $p \geq N$). In this case the outlier does not appear in both matrices simultaneously and the behavior of the detection statistic $\mathcal{D}_{n,p,q}$ is easy to describe. Indeed, the statistic starts to increase at the moment when the outlier appears in the test matrix for the first time, i.e. around $\tau = n + q + M - 1$ (see Fig. 1) and grows until the outlier appears in the matrix the same number of times, namely $\tilde{M} = \min\{M, Q\}$. Then the statistic becomes stable during $\tilde{Q} = \max\{M, Q\}$ steps and then it starts to decrease (since the number of appearances of the outlier in the test matrix decreases). Thus,

- The statistic increases for $(\tau - q - \tilde{M} + 1) \leq n \leq (\tau - q)$;
- It stabilizes at a high level for $(\tau - q) \leq n \leq (\tau - p - \tilde{M})$;
- The statistic decreases for $(\tau - p - \tilde{M}) \leq n \leq (\tau - p - 1)$.

Expectation of $\tilde{\mathcal{D}}_{n,p,q}$ (without averaging) at iteration n is $E\tilde{\mathcal{D}}_{n,p,q} = 1 + ((k(A^2 - 1))/MQ)$, where $1 \leq k \leq \tilde{Q}$ is the number of times the outlier appears in the test matrix at iteration n, A is the value of the outlier (assuming that the signal has being subtracted). In particular, if $n \in [\tau - q, \tau - p - \tilde{M}]$ then $k = \tilde{Q}$ and therefore the expected value of the detection statistics at the highest level is $E\tilde{\mathcal{D}}_{n,p,q} = 1 + (A^2 - 1)/\tilde{M}$.

An example of a time series with outlier is shown in Fig. 4. In this figure the sequence $\{e_i\}$ consists of i.i.d. normal random variables, i.e., $e_i \sim N(0, 1)$ for each $i = 1 \ldots 200$. The outlier is at point $i = 100$ and its value is $A = 8$. The detection statistic $\tilde{\mathcal{D}}_{n,p,q}$ (see Fig. 5) is obtained by using the algorithm with parameters $N = 50$, $M = 25$, $p = 26$, $K = N - M + 1 = 26$ and $Q = 25$.

6. NUMERICAL APPROXIMATIONS FOR THE ERROR PROBABILITIES

6.1. Performance Characteristics

In this section results of simulations are reported that are used to assess the quality of different approximations and compare the proposed algorithm against some classical change-point detection algorithms.

As the quality characteristics we consider the error probabilities of type I and II and the corresponding power function. We do not consider standard in sequential change-point detection theory criteria such as expected run length and average time to false detection. The reason is that for the detection statistics like the moving sum (our detection statistic is the moving weighted sum of squares) it is more natural to use the
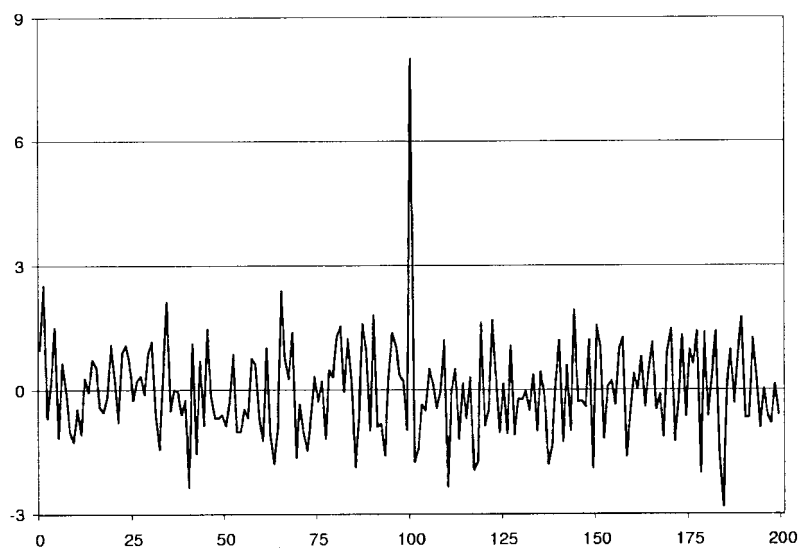
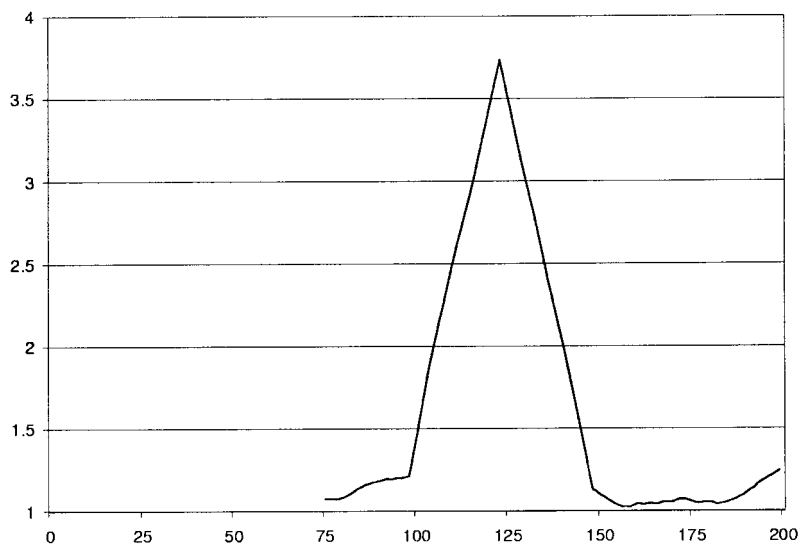***Figure 4.*** Gaussian sequence with an outlier.

Figure 5. The detection statistic $\tilde{\mathcal{D}}_{n,p,q}$, $N = 50$, $M = 25$, $p = 26$, $q = 51$.

approach which can be called "reliable detection" (see Bakhache and Nikiforov (2000)). In this approach two time interval lengths, say T_0 and T_1, are fixed and the following two probabilities are considered:

$$\sup_{k \geq 1} \Pr\{S_n \geq h \text{ for at least one } n = k+1, \ldots, k + T_0 \,|\, \mathbf{H_0}\} \tag{11}$$

$$\sup_{\tau \geq 1} \Pr\{S_n < h \text{ for all } n = \tau + 1, \ldots, \tau + T_1 \,|\, \mathbf{H_1}(\tau)\} \tag{12}$$

Here S_n is the detection statistic (like moving sum), h is the threshold, $\mathbf{H_0}$ is the null-hypothesis of no change, $\mathbf{H_1}(\tau)$ denotes the hypothesis that the change occurred at time τ, T_1 is the maximum possible delay in detecting the signal, T_0 is the interval of time where we monitor the probability of false alarm. Supremum in both Eqs. (11) and (12) disappears if the statistics S_n form a stationary sequence under the null hypothesis; this is the case in our study as soon as the change does not happen too soon (we must have time to form the first sum of squares while there is no change).

The probabilities (11) and (12) will be called the error probabilities of type I and type II, respectively. A natural choice of T_0 and T_1 is $T_0 = T_1 = T$, say; we often implicitly assume this while studying the error probabilities of type II (and the corresponding power function). In these cases it is reasonable to assume that T is not very large, respectively to M. In most cases, however, we study only the error probabilities of type I (which defines the significance level of the test) and can assume that $T = T_0$ is much larger than M.

Formally, if the null hypothesis is $\mathbf{H_0}$ and the alternative is $\mathbf{H_1}$, then the error probability of type I (the probability of false alarm) is

$$\alpha_I = \Pr\{\text{reject } \mathbf{H_0} \,|\, \mathbf{H_0}\} = \Pr\{S \geq h \,|\, \mathbf{H_0}\}, \tag{13}$$

and the error probability of type II (the probability of missing the signal) is

$$\alpha_{II} = \Pr\{\text{accept } \mathbf{H_0} \,|\, \mathbf{H_1}\} = \Pr\{S < h \,|\, \mathbf{H_1}\}, \tag{14}$$

where S is a detection statistic and the test here is defined as $S \lessgtr_{\mathbf{H_1}}^{\mathbf{H_0}} h$ with $S < h$ meaning that $\mathbf{H_0}$ is accepted. In many cases the hypothesis $\mathbf{H_0}$ is simple but the hypothesis $\mathbf{H_1}$ is complex and parametrized by some parameter θ (with $\theta = \theta_0$ corresponding to the hypothesis $\mathbf{H_0}$). Then the error probability α_{II} depends on θ, that is, $\alpha_{II} = \alpha_{II}(\theta)$. The error probabilities are usually expressed in the form of the power function

$$\beta(\theta) = 1 - \alpha_{II}(\theta) = \Pr\{S \geq h \,|\, \theta\}, \tag{15}$$

with $\beta(\theta_0) = \alpha_I$.

6.2. Null Hypothesis Model

The underlying assumption of the SSA technique in general and the proposed change-point detection algorithm in particular is the assumption that the initial time series is well approximated by the series z_t, the solution of the finite-difference equation:

$$z_t = a_1 z_{t-1} + \cdots + a_d z_{t-d} \tag{16}$$

of some order d with some coefficients $a_1, \ldots, a_d$, that is, by a process of the form

$$z_t = \sum_k \alpha_k(t) e^{\mu_k t} \sin(2\pi \omega_k t + \varphi_k). \tag{17}$$

(here $\alpha_k(t)$ are polynomials in t, μ_k, ω_k and φ_k are arbitrary parameters) with a small number of terms. That is, we assume that

$$x_t = z_t + e_t \tag{18}$$

where e_t is a noise process and z_t satisfies the finite-difference equation

$$z_t = a_1 z_{t-1} + \cdots + a_d z_{t-M} \tag{19}$$

with some coefficients $a_1, \ldots, a_M$ and certain initial conditions. (Note that some of the coefficients a_i can be zero and thus the order of the finite-difference Eq. (19) can be smaller than M.) The noise can be either random or deterministic, but it must have the property that its approximation by the solutions of the finite-difference equations is poor (white noise certainly satisfies this assumption).

Application of SSA with lag M at time intervals $[n+1, n+N]$ approximately recovers the model (18). That is, we get

$$x_t = z_t^{(n)} + e_t^{(n)}, \tag{20}$$

where $z_t^{(n)}$ is the SSA approximation for z_t, the solution of Eq. (19).

Asymptotically, when $N \to \infty$, $M \to \infty$, and the noise e_t is an ergodic random process with finite variance, we obtain that these two processes are weakly asymptotically separable in the sense that **w**-correlations defined in Eq. (9) tend to 0 (see Goljandina, et al. (2001), Chapter 6 and Corollary 6.1; Danilov and Zhigljavsky (1997), p. 221). In computing the threshold h we assume the following null-hypothesis:

1. The model Eq. (18) is valid and there is no change in parameters of the Eq. (19),
2. $z_t^{(n)} = z_t$ for all n and t,

3. Either M or $Q = q - p$ tend to infinity,
4. $e_t = e_t^n$ is a sequence of i.i.d.r.v., the variance σ^2 of e_t can be unknown.

Note, that from theoretical point of view, values of p and q are not important. The important quantity is the difference $Q = q - p$. For simplicity let us assume $M \geq Q$ (if $M \leq Q$ then all the formulas can be obtained substituting $M \leftrightarrow Q$).

The above assumptions imply that in the change-point detection algorithm we have at iteration n

$$\mathcal{D}_{n,I,p,q} = \sum_t w_{M,n+p,n+q}(t)e_t^2, \tag{21}$$

where (for $Q \leq M$)

$$w_{M,p,q}(t) = \begin{cases} t - p & \text{for } p < t \leq q, \\ Q & \text{for } q < t \leq M + p, \\ q + M - t & \text{for } M + p < t \leq M + q, \\ 0 & \text{otherwise.} \end{cases} \tag{22}$$

The form of the weight function $w_{M,p,q}(t)$ is caused by the structure of the trajectory matrix Eq. (5), where x_1 appears once, x_2 – twice, etc.

Obviously, Eq. (21) is a quadratic form $e^T B e$, where $e = (e_1, e_2, \ldots, e_N)^T$ and $B = B(M, n, p, q)$ is a diagonal matrix with diagonal elements $B_{tt} = w_{M,n+p,n+q}(t)$.

In Moskvina (2001; 2000) the distribution of the quadratic form $\mathcal{D}_{n,I,p,q}$ is studied and the moments of this quadratic form, $E\mathcal{D}_{n,I,p,q}$ and $\text{var}(\mathcal{D}_{n,I,p,q})$, are calculated for the general case when the averaging of the base matrix is possibly performed. In the simple case of no averaging we have

$$E(e^T B e) = \sigma^2 \text{tr} B = MQ\sigma^2, \tag{23}$$

$$\text{var}(e^T B e) = 2\sigma^4 \text{tr} B^2 = 2\sigma^4 \frac{1}{3} Q(3MQ + 1 - Q^2) \quad (Q \leq M). \tag{24}$$

Standardizing the random variable $\mathcal{D}_{n,I,p,q} = e^T B e$ and taking into account its asymptotic normality (when M or Q tend to infinity), which is a consequence of (iii), we get asymptotically

$$\frac{\mathcal{D}_{n,I,p,q} - E\mathcal{D}_{n,I,p,q}}{\sqrt{\text{var}(\mathcal{D}_{n,l,p,q})}} \sim N(0,1). \tag{25}$$

6.3. Numerical Approximation of the Error Probabilities of Type I

In this section we provide results of simulations that illustrate the quality of approximations of the error probabilities of first and second kind for the moving squared distances by the corresponding probabilities for moving sums of normal random variables. In each reported case we performed 10 000 simulations of the corresponding statistic; the results presented are the values of the corresponding proportions.

Sum of Squared Distances

According to the results of the previous section, one of the main detection statistics in the SSA-based algorithm can be represented as a moving sum of the squared distances to the l-dimensional subspace that describes the main structure of the series; it can often be associated with the signal. Under the assumption that the signal is well described by this subspace and there is no change in the structure of the series, the residuals can often be considered as i.i.d.r.v. Thus, the moving squared distance is considered as a moving weighted sum of squares of these random variables.

Let us study the change-point detection algorithm under the null hypothesis (see the previous section). Assume that the noise $\{e_t\}$ consists of i.i.d.r.v. with $Ee_t = 0$ and $\mathrm{var}(e_t) = \sigma^2$. We shall also assume that M is large enough so that υ_n, which is a consistent estimate of $E\upsilon_n = E\mathcal{D}_{n,p,q} = MQ\sigma^2$, is close enough to $MQ\sigma^2$. Using the Slutski theorem (see Rao (1973), p.122) we then can replace σ^2 by υ_n/MQ (this does not affect the asymptotic distribution of $\mathcal{D}_{n,p,q}$). Thus we can assume $\sigma^2 = 1$, i.e. that e_t are normal i.i.d. random variables $N(0,1)$.

Studying the behavior of $\mathcal{D}_{n,p,q}$ is obviously equivalent to studying the behaviour of the sequence of random variables defined by

$$\xi_n = \frac{\mathcal{D}_{n,p,q} - E\mathcal{D}_{n,p,q}}{\sqrt{\mathrm{var}(\mathcal{D}_{n,p,q})}} \quad (n = 0, \ldots, T - N). \tag{26}$$

Indeed, the event that ξ_n reaches the threshold h (i.e. $\xi_n \geq h$ for some n) is equivalent to

$$\mathcal{D}_{n,p,q} \geq h\sqrt{\mathrm{var}(\mathcal{D}_{n,p,q})} + E\mathcal{D}_{n,p,q} = H.$$

For simulations we use the standardized random quadratic form Eq. (26). Using formulae (23) and (24) we get $E\mathcal{D}_{n,p,q} = MQ$, and $\mathrm{var}(\mathcal{D}_{n,p,q}) = 2Q(3MQ - Q^2 + 1)/3$. Thus, we simulate the random variables

$$\xi_n = \frac{\sum_{i=1}^{M+Q-1} w_{M,p+n,q+n}(i)e_i^2 - MQ}{\sqrt{2Q(3MQ - Q^2 + 1)/3}} \,.$$

The behavior of the error probability of type I as a function of T and h for the sequence of r.v. ξ_n $(n = 0, \ldots, T - N)$ is shown in Fig. 6.

Sum of Normal Random Variables

Consider the sequence of random variables

$$\xi'_n = \frac{\mathcal{D}'_{n,p,q} - E\mathcal{D}'_{n,p,q}}{\sqrt{\mathrm{var}(\mathcal{D}'_{n,p,q})}} \,,$$

where $\mathcal{D}'_{n,p,q} = \sum_t w_{M,n+p,n+q}(t)e_t$ $(t = 1, \ldots, T)$ and e_t are normal i.i.d. $N(0, \sigma^2)$. The mean and variance of the random variable ξ'_n are

$$E\xi_n = E\xi'_n = 0 \quad \text{and} \quad \mathrm{var}(\xi_n) = \mathrm{var}(\xi'_n) = 1.$$
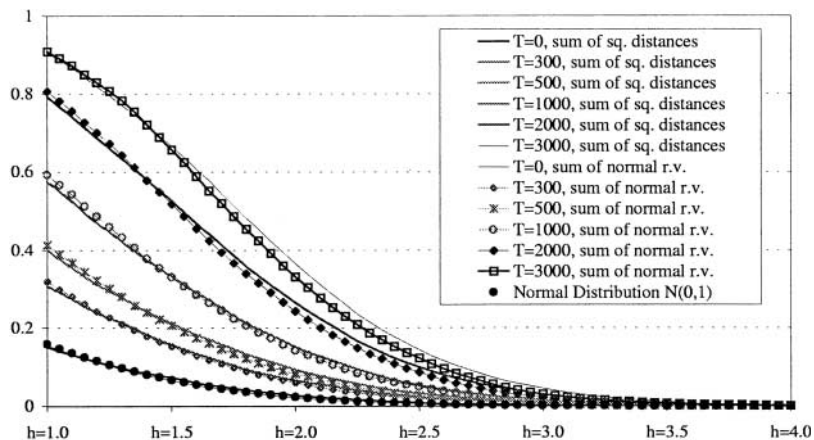

Figure 6. The error probabilities of type I. Sum of squared distances and sum of normal r.v. $M = 300$, $Q = 300$.

Straightforward calculations give

$$E\mathcal{D}'_{n,p,q} = 0 \quad \text{and} \quad \text{var}(\mathcal{D}'_{n,p,q}) = \frac{\sigma^2 Q}{3}(3MQ - Q^2 + 1).$$

Therefore, we simulate the random variables

$$\xi'_n = \frac{\sum_{i=1}^{M+Q-1} w_{M,p,q}(t)e_i}{\sqrt{Q(3MQ - Q^2 + 1)/3}},$$

where $e_i \sim N(0,1)$ for any i.

The error probabilities α_I of type I (see Eq. (13)) (i.e. the probability of the event that the algorithm indicates a change when there is not one) is illustrated in Fig. 6 for $M = Q = 300$. This figure shows that the error probabilities of type I for sums of distances and squared distances are close for reasonably large values of parameters M and Q.

7. NUMERICAL APPROXIMATION FOR THE POWER FUNCTION AND COMPARISON OF ALGORITHMS

In first part of this section we consider the problem of testing between two hypotheses:

$$\mathbf{H_0}: \quad \theta = \theta_0, \quad \text{for any} \quad t \in [0, T] \tag{27}$$

and the alternative of the change in mean:

$$\mathbf{H_1}: \begin{cases} \theta = \theta_0, & 0 \le t \le \tau, \\ \theta = \theta_1, & \tau < t \le T. \end{cases} \tag{28}$$

7.1. SSA-Based Algorithm

Let us consider a Gaussian sequence with change in mean and constant variance $\sigma^2 = 1$. Fig. 7 show the behaviour of the power function $\beta(a)$ (see Eq. (15)) for the proposed SSA-based change-point detection algorithm with different sets of parameters. The range of change in mean is from $a = 0$ till $a = 0.7$ is indicated by a in the legend of each graph. Note, that the black lowest curve corresponding to $a = 0$ is α_I, the error probability of type I.

To compare the power function of the test $\beta(a) = 1 - \alpha_{II}$ for different parameters M and Q let us fix the error probability of type I to be
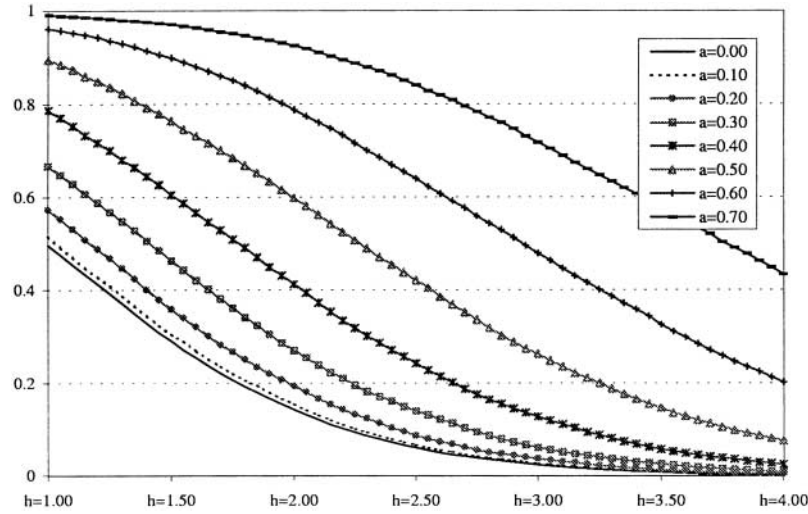
Figure 7. Power functions of the SSA-based algorithm: change in mean, $M = 90$, $Q = 1$.

$\alpha_I = 0.100$ and $T' = T - N = 100$, where $N = M + Q - 1$. Note, that we need the first N observations to build the quadratic form

$$\mathcal{D}_{0,0,Q} = \sum_{i=1}^{N} w_{M,O,Q}(i)e_i^2$$

(see Eq. (21)) on the first step and then T' steps to test the algorithm. Table 1 demonstrates probabilities of reaching the threshold h for the sum of squares of the normal r.v. under the condition $\alpha_I = 0.100$ with different M and Q. The largest value of the power function is reached at $M = 90$ and $Q = 1$.

7.2. CUSUM Algorithm

Cumulative sum detection statistic (CUSUM) for Gaussian sequence is

$$g_k = S_k - \min_{1 \le i \le k} S_i \underset{\mathbf{H_1}}{\overset{\mathbf{H_0}}{\lessgtr}} h,$$

where

$$S_k = \frac{\Delta}{\sigma^2} \sum_{i=1}^{k} \left(x_i - \theta_0 - \frac{\Delta}{2} \right), \quad \Delta = \theta_1 - \theta_0.$$

Table 1. Probabilities of reaching the threshold h for the SSA-based algorithm: change in mean.

M, Q	50, 1	80, 1	80, 10	90, 1	90, 10	100, 1	100, 10	50, 50	100, 100
$a = 0.1$	0.114	0.108	0.107	0.108	0.109	0.110	0.108	0.107	0.103
$a = 0.2$	0.144	0.141	0.133	0.140	0.136	0.139	0.135	0.132	0.118
$a = 0.3$	0.213	0.209	0.201	0.206	0.196	0.204	0.197	0.187	0.147
$a = 0.4$	0.337	0.329	0.322	0.330	0.315	0.326	0.311	0.293	0.202
$a = 0.5$	0.505	0.513	0.508	0.520	0.495	0.509	0.491	0.450	0.291
$a = 0.6$	0.702	0.722	0.716	0.728	0.707	0.720	0.702	0.656	0.436
$a = 0.7$	0.867	0.887	0.884	0.892	0.881	0.888	0.875	0.834	0.619
$a = 0.8$	0.958	0.971	0.969	0.972	0.969	0.971	0.966	0.948	0.788
$a = 0.9$	0.991	0.995	0.995	0.996	0.995	0.994	0.995	0.988	0.914
$a = 1.0$	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.998	0.974

This statistic depends on parameters θ_0 and θ_1, which are usually unknown. Let us consider the following version of the CUSUM algorithm (for details see Basseville and Nikiforov (1993))

$$g'_k = S'_k - \min_{1 \leq i \leq k} S'_i \underset{\mathbf{H_1}}{\overset{\mathbf{H_0}}{\lessgtr}} h', \quad \text{where} \quad S'_k = \frac{1}{\sigma^2} \sum_{i=1}^{k} x_i. \tag{29}$$

To compare our and the CUSUM algorithm (29) consider the probability of reaching the threshold h by both algorithms without signal (the error probability of type I, see Eq. (13)) and with signal (the power function (15)) at time $T = 100$. The largest value of the power function for our algorithm is with $M = 90$, $Q = 1$. Let us again fix the error probability of type I to be $\alpha_I = 0.100$ for $T = 100$. Using the simulation we obtain that the value of the threshold is $h = 18$.

Table 2 shows the probabilities of reaching this threshold with different T for the value of signal $a = 0.0, 0.1, \ldots, 1.0$; see also Fig. 8. The bold font numbers show approximate matches with the column $M = 90$, $Q = 1$ of Table 1. As we can see, the CUSUM algorithm is about three times better than the SSA-based algorithm: indeed, the probability which we obtain with our algorithm for 100 steps can be reached by the CUSUM algorithm for ~ 36 steps.

The power functions for these two tests are displayed in Fig. 9. The power function of the proposed (SSA-based) algorithm is presented with parameters $M = 90$ and $Q = 1$, where the power function is maximum (see Table 1). The power function of the CUSUM test shown in the figure is presented for $T = 36$. The thresholds for the tests have been chosen so that the values of the power functions are close (~ 0.1) for $a = 0$.

Table 2. Probabilities of reaching the threshold $h = 18$ for the CUSUM algorithm: change in mean.

	$T = 35$	$T = 36$	$T = 37$	$T = 40$	$T = 45$	$T = 50$
$a = 0.1$	0.012	0.014	0.016	0.024	0.039	0.061
$a = 0.2$	0.044	0.047	0.059	0.085	**0.129**	0.186
$a = 0.3$	0.127	0.143	0.158	**0.216**	0.314	0.419
$a = 0.4$	0.283	0.316	**0.341**	0.430	0.568	0.683
$a = 0.5$	0.505	**0.538**	0.571	0.667	0.794	0.878
$a = 0.6$	**0.716**	0.749	0.779	0.853	0.928	0.967
$a = 0.7$	0.873	**0.896**	0.914	0.954	0.982	0.993
$a = 0.8$	0.960	0.968	**0.974**	0.989	0.996	0.998
$a = 0.9$	0.989	0.992	**0.994**	0.998	0.999	0.999
$a = 1.0$	0.998	0.998	0.999	0.999	0.999	1
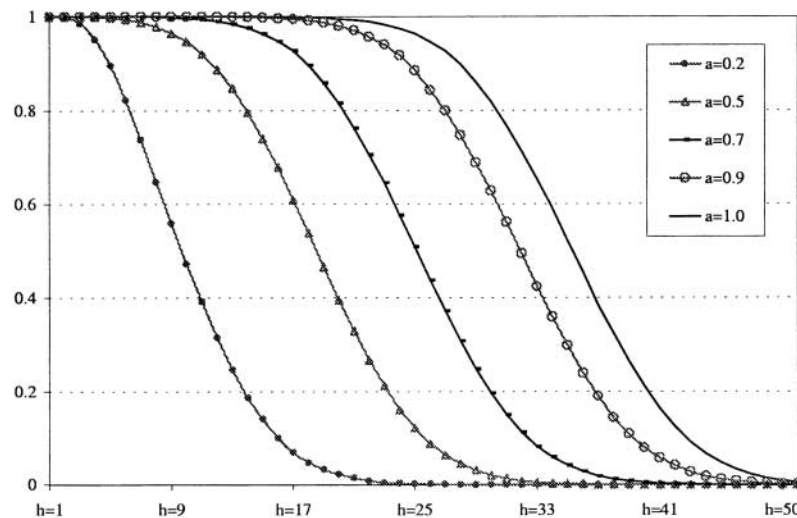
Figure 8. Power functions for the CUSUM test: change in mean, $T = 36$.

7.3. Error Probabilities When the Change Is in Variance

Let us consider a Gaussian sequence with zero mean and change in variance σ^2. An example of the behaviour of the power function $\beta(a)$ for the proposed SSA-based change-point detection algorithm is displayed in Fig. 10. The range of change of the standard deviation is from 1.0 till 1.5,
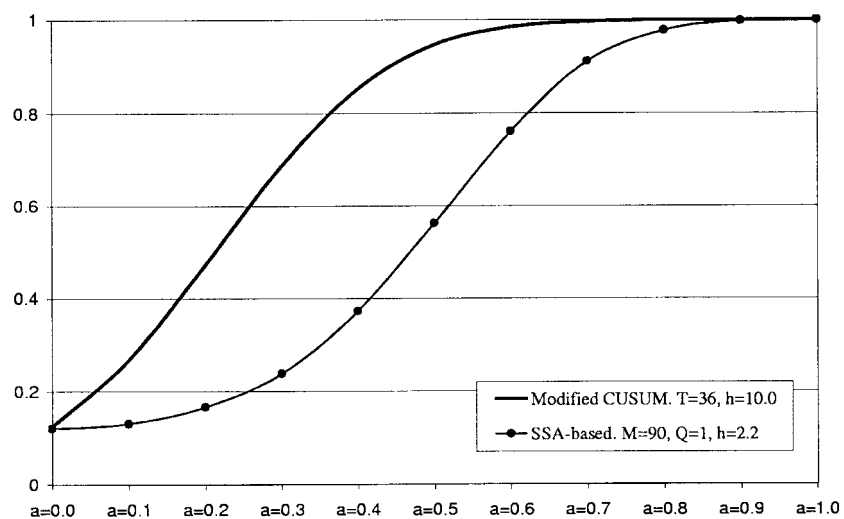
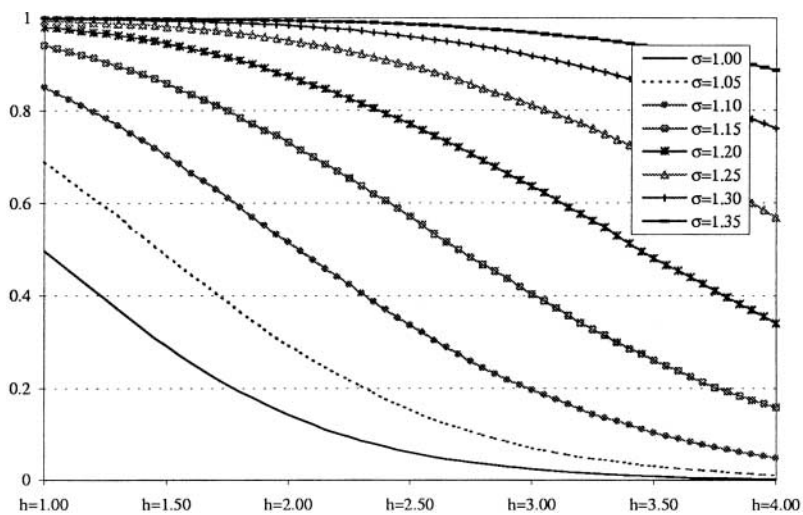***Figure 9.*** Power functions of the SSA-based and CUSUM tests as a function of *a*.

Figure 10. Power functions for the SSA-based algorithm: change in variance, $M = 90$, $Q = 1$.

Table 3. Probabilities of reaching the threshold h for the SSA-based algorithm: change in variance.

M, Q	80, 1	80, 10	90, 1	90, 10	100, 1	100, 10	100, 100
$\sigma = 1.05$	0.236	0.227	0.232	0.223	0.227	0.219	0.150
$\sigma = 1.10$	0.441	0.425	0.441	0.427	0.437	0.425	0.240
$\sigma = 1.15$	0.665	0.651	0.670	0.653	0.655	0.647	0.379
$\sigma = 1.20$	0.837	0.826	0.837	0.827	0.826	0.817	0.535
$\sigma = 1.25$	0.933	0.928	0.934	0.927	0.931	0.918	0.695
$\sigma = 1.30$	0.974	0.973	0.977	0.974	0.979	0.968	0.812
$\sigma = 1.35$	0.992	0.992	0.992	0.991	0.992	0.990	0.896
$\sigma = 1.40$	0.998	0.998	0.998	0.997	0.998	0.997	0.951
$\sigma = 1.45$	0.999	0.999	0.999	0.999	0.999	0.999	0.978
$\sigma = 1.50$	1.000	0.999	0.999	0.999	0.999	0.999	0.991

Table 4. Probabilities of reaching the threshold $h = 18$ for the CUSUM algorithm: change in variance.

	$T = 50$	$T = 100$	$T = 200$	$T = 300$	$T = 400$	$T = 500$	$T = 600$
$\sigma = 1.05$	0.019	0.129	0.391	0.580	0.709	0.798	0.859
$\sigma = 1.10$	0.026	0.154	0.429	0.620	0.746	0.830	0.886
$\sigma = 1.15$	0.033	0.178	0.466	0.657	0.776	0.854	0.908
$\sigma = 1.20$	0.044	0.203	0.501	0.693	0.807	0.877	0.929
$\sigma = 1.25$	0.055	0.231	0.540	0.726	0.834	0.899	0.943
$\sigma = 1.30$	0.067	0.259	0.571	0.756	0.858	0.918	0.955
$\sigma = 1.35$	0.080	0.289	0.603	0.785	0.878	0.934	0.965
$\sigma = 1.40$	0.096	0.319	0.633	0.808	0.895	0.946	0.973
$\sigma = 1.45$	0.109	0.347	0.664	0.834	0.912	0.957	0.978
$\sigma = 1.50$	0.124	0.373	0.695	0.857	0.927	0.966	0.984

it is indicated by σ in the legend of each graph. The black lowest curve corresponding to $\sigma = 1.0$ is α_I, the error probability of type I.

Table 4 shows the probabilities of reaching this threshold with different T for the value of standard deviations $\sigma = 1.0, 0.05, \ldots, 1.5$; see also Fig. 11. Let us again find numbers in this table approximate matching with the column $M = 90$, $Q = 1$ of Table 3. As one can see, the CUSUM (29) algorithm is now about six times worse than the SSA-based algorithm: indeed, the probability which we obtain with our algorithm for 100 steps can be reached by the CUSUM algorithm for ~ 600 steps. As it was mentioned in Example 2, the likelihood ratio test
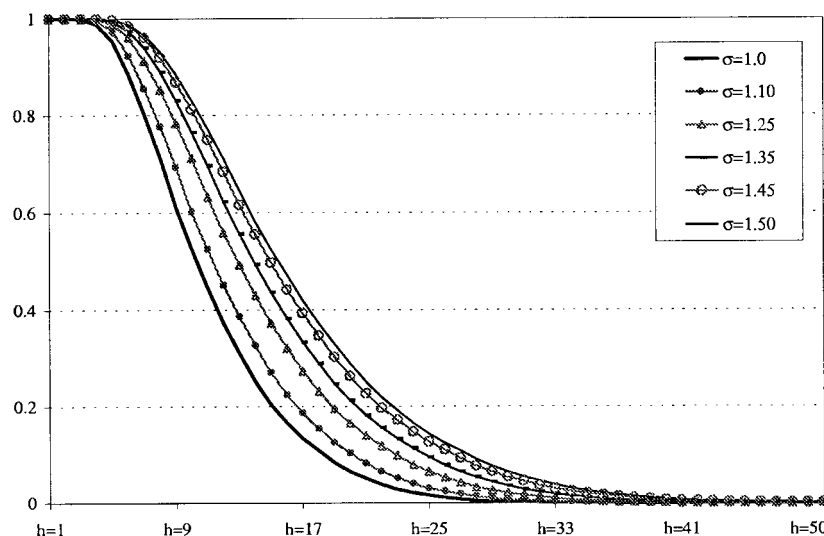
Figure 11. Power functions for the CUSUM test: change in variance, $T = 100$.

for this problem is exactly the moving sum of squares; so the result is not surprising.

8. THE ERROR PROBABILITIES AT FIXED ITERATION

Denote

$$S = \sum_{i=1}^{M} x_i^2, \qquad S' = \sum_{i=1}^{M} x_i,$$

and compare S and S' when x_i are i.i.d. normal random variables ($x_i \sim N(a, \sigma^2)$). For fixed iteration number S is the detection statistic appearing as a particular case of SSA-based algorithm for $Q = 1$:

$$\mathcal{D}_{n, p, p+1} = \sum_{i=1}^{M} x_{n+i}^2$$

(in this section we drop the group I out of the subscript $\mathcal{D}$ since we focus here on the issues that are irrelevant to I). $S' = \sum_{i=1}^{M} x_i$ is related to the CUSUM test Eq. (29).

The moments of the random variables S and S' can be easily computed:

$$ES' = Ma, \text{var}(S') = M\sigma^2; \quad ES = M(a^2 + \sigma^2), \quad \text{var}(S) = 2M\sigma^2(2a^2 + \sigma^2).$$

In this section we compare the quality of the tests based on the statistics S and S' evaluated at a fixed time moment n for the change-point model (27)–(28) with normal i.i.d. noise.

8.1. Testing a Simple Hypothesis Against a Simple Alternative

Let us consider a simple auxiliary problem of testing between two simple hypotheses. Specifically, consider the problem of testing the hypothesis

$$\mathbf{H_0}: \quad \xi = \xi_0 \sim \varphi_0\left(\frac{x}{\sigma_0}\right),$$

(here we have $E\xi_0 = 0$, $\text{var}\,\xi_0 = \sigma_0^2$, $\xi_0 = \eta_0\sigma_0$ and $\eta_0 \sim \varphi_0$) against the alternative

$$\mathbf{H_1}: \quad \xi = \xi_1 \sim \varphi_1\left(\frac{x - \mu}{\sigma_1}\right)$$

(where $E\xi_1 = \mu \geq 0$, $\text{var}\,\xi_1 = \sigma_1^2$, $\xi_1 = \mu + \eta_1\sigma_1$, $\eta_1 \sim \varphi_1$). Note that $E\eta_0 = E\eta_1 = 0$ and $\text{var}\,\eta_0 = \text{var}\,\eta_1 = 1$.

The criterion and the error probability of type I (the probability of false alarm) are

$$\xi \lessgtr_{\mathbf{H_1}}^{\mathbf{H_0}} h = \sigma_0\, t_\alpha, \tag{30}$$

$$
\begin{aligned}
\alpha_I &= \Pr\{\text{reject } \mathbf{H_0} \,|\, \mathbf{H_0}\} \\
&= \Pr\{\xi_0 \geq \sigma_0 t_\alpha\} = 1 - \Pr\{\eta_0 < t_\alpha\} = 1 - F_0(t_\alpha) = \alpha,
\end{aligned}
$$

where t_α is the $(1 - \alpha)$-quantile of the null distribution:

$$F_0(t_\alpha) = \int_{-\infty}^{t_\alpha} \varphi_0(t)\, dt = 1 - \alpha.$$

The error probability of type II, the probability of missing the signal, is

$$\alpha_{II} = \Pr\{\text{accept } \mathbf{H_0} \,|\, \mathbf{H_1}\} = \Pr\{\xi_1 < \sigma_0 t_\alpha\} = \Pr\left\{\eta_1 < \frac{\sigma_0 t_\alpha - \mu}{\sigma_1}\right\}.$$

Thus,

$$\alpha_{II} = F_1\left(\frac{\sigma_0 t_\alpha - \mu}{\sigma_1}\right), \tag{31}$$

where $F_1(t)$ is the c.d.f. related to $\varphi_1(t)$. The power function of the criterion is

$$\beta = \beta(\mu, \sigma_1) = 1 - \alpha_{II} = 1 - F_1\left(\frac{\sigma_0 t_\alpha - \mu}{\sigma_1}\right).$$

8.2. Approximation of the Error Probabilities
(Change in Mean)

Let us apply the formulae of the previous section to the statistic S'. Here σ_0 and σ_1 are $\sigma_0 = \sigma_1 = \sqrt{M}\sigma$, $\mu = Ma$ and the distributions are normal. Thus, we have the hypotheses:

$$\mathbf{H_0} : S' \sim N(0, M\sigma^2),$$
$$\mathbf{H_1} : S' \sim N(Ma, M\sigma^2).$$

The criterion is

$$S' \underset{\mathbf{H_1}}{\overset{\mathbf{H_0}}{\lessgtr}} h' = \sqrt{M}\sigma\, t_\alpha,$$

where t_α is the $(1 - \alpha)$-quantile of the standard normal distribution:

$$\Phi(t_\alpha) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t_\alpha} e^{-(t^2/2)}\, dt = 1 - \alpha, \tag{32}$$

and $\Phi(x)$ is the c.d.f. of the standard normal distribution. The error probability of type II is and the power function of the test are

$$\alpha_{II} = \Phi\left(t_\alpha - \frac{a\sqrt{M}}{\sigma}\right), \quad \beta = \beta(a) = 1 - \Phi\left(t_\alpha - \frac{a\sqrt{M}}{\sigma}\right). \tag{33}$$

Consider the statistic

$$\xi = \frac{1}{\sigma^2} S - M = \sum_{i=1}^{M} \left(\frac{x_i^2}{\sigma^2}\right) - M.$$

Asymptotically, when $M \to \infty$, for the statistic $\xi = (1/\sigma^2)S - M$ we have the hypotheses:

$\mathbf{H_0}$: $\xi \sim N(0,\ 2M)$,
$\mathbf{H_1}$: $\xi \sim N\left(\frac{Ma^2}{\sigma^2},\ 2M\left(\frac{2a^2}{\sigma^2}+1\right)\right)$.

The criterion is

$$\xi \lessgtr_{\mathbf{H_1}}^{\mathbf{H_0}} h = \sqrt{2M}\, t_\alpha,$$

where t_α is the $(1-\alpha)$-quantile of the standard normal distribution $\Phi(x)$ (see Eq. (32)). Thus, the error probability of type II (using formula (31)) is:

$$\alpha_{II} = \Phi\left(\frac{\sqrt{2}\sigma^2 t_\alpha - \sqrt{M}a^2}{\sqrt{4a^2\sigma^2 + 2\sigma^4}}\right)$$

and the power function is:

$$\beta(a) = 1 - \Phi\left(\frac{\sqrt{2}\sigma^2 t_\alpha - \sqrt{M}a^2}{\sqrt{4a^2\sigma^2 + 2\sigma^4}}\right). \tag{34}$$

Figure 12 illustrates the behaviour of the power functions (33) and (34) for these two tests with $M = 100$, $\sigma = 1$ and $t_\alpha = 1.2815$, that corresponds to the 0.9-quantile of the standard normal distribution.

Clearly, the S'-based test is the best (it is the optimal likelihood ratio test). The power reduces quite visibly, see Fig. 12. Of course, application of the test based on S further reduces the power: the S-based test is approximately 3 times worse than the optimal S'-based test; that is, one needs to have about $3k$ observations to achieve, for the fixed significance level, the same power with the S-based test as we get for k observations for S'-based test. These conclusions are in agreement with the conclusions of Secs. 7.1–7.2, where the comparisons concerning the same tests, but considered as sequential ones, are made.

8.3. Approximation of the Error Probabilities (Change in Variance)

Consider now the situation when $a = 0$ for both hypotheses but $\sigma_0 = 1$ and $\sigma_1 = \sigma$. Then for the statistic S' we have the hypotheses:

$\mathbf{H_0}$: $S' \sim N(0, M)$,
$\mathbf{H_1}$: $S' \sim N\left(0, M\sigma^2\right)$.
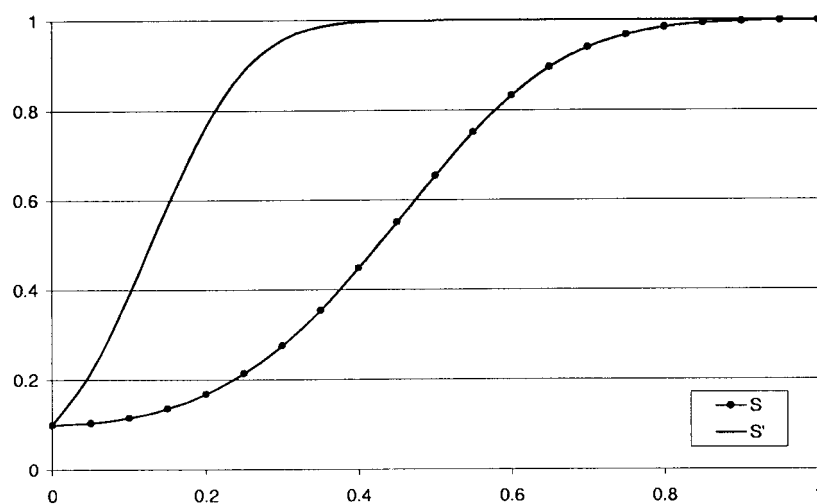
Figure 12. Powers of the tests based on S and S' as functions of a (change in mean).

The criterion is

$$S' \lessgtr^{\mathbf{H_0}}_{\mathbf{H_1}} h' = \sqrt{M}\, t_\alpha, \tag{35}$$

where t_α is defined in Eq. (32). The power function of the test is

$$\beta = \beta(\sigma) = 1 - \Phi\left(\frac{t_\alpha}{\sigma}\right). \tag{36}$$

For a two-sided test $|S'| \lessgtr^{\mathbf{H_0}}_{\mathbf{H_1}} h'' = \sqrt{M}\, t_{\alpha/2}$ the power function is

$$\beta = \beta(\sigma) = 2\Phi\left(\frac{t_{\alpha/2}}{\sigma}\right). \tag{37}$$

Asymptotically (as $M \to \infty$), for the statistic $\xi = S - M$ the hypotheses are:

$$\mathbf{H_0} : \xi \sim N(0,\ 2M),$$
$$\mathbf{H_1} : \xi \sim N\big(M(\sigma^2 - 1),\ 2M\sigma^4\big).$$

Using formulae (30)–(31) we get the criterion

$$\xi \lessgtr^{\mathbf{H_0}}_{\mathbf{H_1}} h = \sqrt{2M}\, t_\alpha;$$
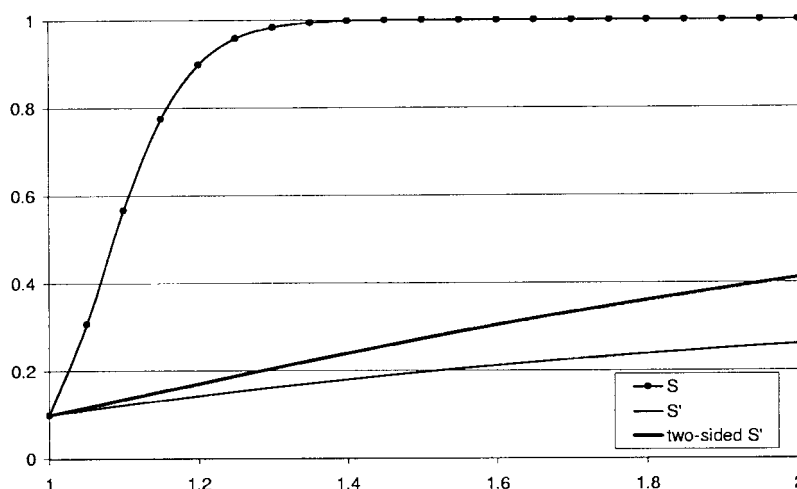
Figure 13. Powers of the tests based on S and S' as functions of σ (change in variance).

the power function is

$$\beta(\sigma) = 1 - \Phi\left(\frac{t_\alpha}{\sigma^2} - \sqrt{\frac{M}{2}}\frac{\sigma^2 - 1}{\sigma^2}\right). \tag{38}$$

Figure 13 illustrates the behaviour of the power functions (36)–(38) for these tests with $M = 100$ and $\alpha = 0.1$. The power of the one-sided test Eq. (35) converges to 0.5. The two-sided test reaches the value 1 but very slowly. In contrast, the power of the test based on S in this case reaches the maximum value 1 quite fast. The numerical results presented in Sec. 7.3 are in full agreement with the theoretical results here.

REFERENCES

Bakhache, B., Nikiforov, I. (2000). Reliable detection of faults in measurement systems. *International Journal of Adaptive Control and Signal Processing* 14:683–700.

Basseville, M., Nikiforov, I. V. (1993). *Detection of Abrupt Changes: Theory and Applications*, Englewood Cliffs, New Jersey: Prentice Hall.

Box, G. E. P., Jenkins, G. M. (1970). *Time series analysis*, *Forecasting and Control*. San-Francisco: Holden-Day.

Broomhead, D. S., King, G. P. (1986). Extracting qualitative dynamics from experimental data. *Physica. D* 20:217–236.

Danilov, D., Zhigljavsky, A. eds. (1997). *Principal Components of Time Series*: "*Caterpillar*" *Method*: St. Petersburg University, [in Russian].

Elsner, J., Tsonis, A. (1996). *Singular Spectrum Analysis*: *a new tool in time series analysis*. New York: Plenum Press.

Goljandina, N., Nekrutkin, V., Zhigljavsky, A. (2001). *Analysis of Time Series Structure*: *SSA and related techniques*. London: Chapman and Hall.

Moskvina, V. (2000). Distribution of random quadratic forms arising in singular-spectrum analysis. *Mathematical Communications* 5:161–171.

Moskvina, V. (2001). Application of the Singular-Spectrum Analysis for Change-Point Detection in Time Series. Ph.D. thesis, School of Mathematics, Cardiff: Cardiff University.

Moskvina, V., Zhigljavsky, A. (2000). Application of singular-spectrum analysis to change-point detection in time series. *Journal of Time Series Analysis*, [*submitted*].

Rao, C. R. (1973). *Linear statistical inference and its applications*. N.Y: Wiley & Sons.

Vautard, R., Yiou, P., Ghil, M. (1992). Singular-spectrum analysis: a toolkit for short, noisy chaotic signals. *Physica. D* 58:95–126.